Vision Res. Vol. 32, No. 5, pp. 983–995, 1992
Printed in Great Britain.

0042-6989/92 $5.00 + 0.00

The Information Capacity of Visual Attention

PREETI VERGHESE,*† DENIS G. PELLI*‡

Received 23 August 1991; in revised form 13 November 1991

Is visual attention mediated by a general-purpose processor with a small data capacity? Such an attentive processor could perform a wide range of transformations upon a small amount of image data. We suggest that this limited capacity corresponds to a fixed amount of information, measured in bits. We measure how much information an observer's attention can handle by measuring how much we can restrict display information without impairing the observer's performance. The attentive visual tasks we study are the detection of a stationary dot in a field of moving dots, and the detection of a static square in a field of flashing squares. Performance of these tasks is perfect up to a critical number of elements (the span of attention) and then falls as the number of elements increases beyond this critical number. The display information required for unimpaired performance in each of these tasks is low; the results indicate that visual attention processes only 30 to 60 bits of display information.

Visual attention Span of attention Critical number Information capacity

INTRODUCTION

The Span of Attention Hypothesis: We conjecture that all attentive visual tasks are performed by an attentive processor. This processor is powerful, able to perform a wide range of perceptual decisions, but has a low data capacity, processing only a small amount of visual information, perhaps 50 bits in each glimpse.

From a glimpse at the night sky we can only reproduce the locations of a handful of stars. Call this limited processing capacity our *span of attention.* Because our span of attention limits the complexity that we perceive, we must serially shift our attention to different areas of the visual scene. We suspect that this fundamental limit of visual perception affects most visual tasks. Our aim is to measure the capacity of this bottleneck in visual processing. Our motivation comes from three lines of work: experiments that distinguish preattentive tasks and attentive tasks, absolute judgment tasks, and studies of the visual requirements of reading and mobility.

Preattentive and attentive tasks

Treisman's and Julesz's extensive work in visual attention supports Neisser's (1967) original conjecture that vision is functionally divided into preattentive and attentive processes. They and others have identified some of the features that segregate effortlessly, in parallel, such as color, orientation, spatial frequency, movement, binocular disparity, and flicker rate (Treisman & Gelade, 1980; Julesz, 1981a,b, 1984; Treisman, 1985; Nakayama & Silverman, 1986). Preattentive tasks are mediated by an array of simple mechanisms, specialized to detect these features. There is no analogous array for visual attention, because it does not seem to be subserved by such special purpose modules. Treisman (1980, 1985) has shown that those targets defined by a combination of features, or characterized by the absence of a feature tend to require visual attention (for exceptions see, e.g. Nakayama & Silverman, 1986; Wolfe, Cave & Franzel, 1989). We think attention is more general purpose and is required to process all visual information not captured by the preattentive mechanism. Visual search in attentive tasks has been described as a series of steps of focal attention with the scan time per element depending on the degree of difficulty of the task. The powerful and sequential nature of attention suggests to us that attention is mediated by a general-purpose processor with a limited data capacity.

We define a *preattentive* task as one in which the probability of detecting the target is independent of the number of distractor elements. For an *attentive* task the probability of detecting the target is inversely proportional to the number of elements in the display.

Absolute judgment tasks

Shannon's definitive work on the information capacity of a communication channel (Shannon & Weaver, 1949) also popularized the concept of the human observer as a communication channel transmitting information about sensory stimuli (also see Hick, 1952; Crossman,

*Institute for Sensory Research, Syracuse University, Syracuse, NY 13244-5290, U.S.A.
†Present address: Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, U.S.A.
‡To whom reprint requests should be addressed.

1953; Miller, 1956). Miller (1956) later reviewed the attempts to measure the channel capacity of the observer in terms of the number of categories into which an observer could divide a single sensory dimension. Miller consolidated data from absolute judgment tasks in vision, hearing and taste, concluding that for identification of stimuli varying along a single sensory dimension, the information capacity of the human observer was the magic number 7 ± 2 categories (about 2 or 3 bits).

Humans fare poorly when compared to physical instruments at judging a single dimension such as pitch or light intensity, but are hard to beat when asked to read text or to recognize familiar faces. Sperling (1960) measured the information transmitted in a glimpse by cuing partial reports for rows of letters. Sperling showed that observers, on average, acquired about 41 bits (9.1 letters), of which they could report only part (but any part) on a given trial. Sperling's estimate of transmitted information—from stimulus to iconic store—is so high partly because Sperling chose a task that people are good at, and partly because Sperling's partial report technique bypassed the bottleneck of the subject's report, which transmitted only 19 bits (4.3 letters) per trial.

Here, we wish to focus on visual attention and the way it constrains the flow of image data into visual perception. We think that attention restricts stimulus information, acting as an early bottleneck in the visual system (Fig. 1), well before the unidimensional categorization limit that Miller (1956) described.

Visual requirements of everyday tasks

The tasks studied in this paper are attentive, in the strict sense defined above. But our approach to studying these tasks arises out of the study of everyday tasks that require visual attention in the ordinary sense of the word. Research into the visual requirements of reading, mobility, face recognition, and sign language has shown that vision must be very severely restricted before performance is impaired (Legge, Pelli, Rubin & Schleske, 1985; Pelli, Legge & Schleste; Pelli, 1986; Ginsberg, 1980; Owsley, Sekuler & Boldt, 1981; Fiorentini, Maffei & Sandini, 1983; Pelli, Goldstein, Tremp & Arend, 1989; Schuchard & Rubin, 1989; Sperling, Landy, Cohen & Pavel, 1985; Pavel, Sperling, Riedl & Vanderbeek, 1987; Pearson, 1983; Pearson & Robinson, 1985). From the bandwidths and window sizes reported in these studies, we estimate* that the number of discrete samples required to represent the visual image are 64 for reading, 23 for mobility, and 64–256 for face recognition.

From the number of samples required to represent the visual image, the number of bits could be derived from the contrast required for each of these tasks if the equivalent noise level for the observers were known

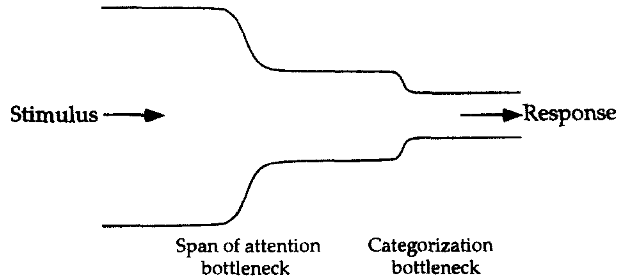

FIGURE 1. A schematic of the constraints on the flow of visual information, in an attentive task, from stimulus to response.

(Pelli, 1990). However, these experiments were not designed with this analysis in mind (pupil size and thus retinal illuminance were not controlled), so we cannot calculate the information in the display required for adequate performance of these tasks. But these experiments do suggest our basic approach: the visual attention span can be described by the number of bits required to represent the raw visual stimulus.

Display information

Attempts in the 1950s to measure the information capacity of the observer by varying the number of different possible stimuli (i.e. source entropy) had limited success. Information theory doesn't care at all about the content of the stimuli; the information transmitted depends only on the joint probabilities of the stimuli and responses. Experiments in memory and coding, however, indicate that the stimulus itself is more important in determining the accuracy of perception and recall than its probability or the ensemble of stimuli that it is drawn from. Miller (1957) reported that for a task requiring the memorization of a list of items, "the variety of the symbols is far less important than the length of the message in controlling what human subjects are able to remember". In other words, Miller (1957) noted that one's ability to faithfully remember a message depends more on the number of "chunks" in the message than on the information content of each chunk.

It is an everyday fact that complicated images are harder to perceive and remember than simple images. For a two-dimensional image made up of dots, the dot number is a useful measure of complexity or "message length". From a glance at the night sky we can recall the arrangement of only a handful of stars. And we can immediately and reliably judge dot number only up to about seven dots (Jevons, 1871; Kaufman Lord, Reese & Volkmann, 1949; Atkinson, Campbell & Francis, 1976; Atkinson, Francis & Campbell, 1976). However, it seems to us that this perceptual complexity limit must be general, applying also to continuous images, e.g. natural scenes, that are not easily specified merely as collections of discrete elements.

We sought to devise a measure of complexity that would be applicable to any image. We reasoned that early vision is probably unable to adapt its coding to any particular experiment's set of stimuli, and would instead have to encode a retinal image in a way that could represent practically any natural image of the same size

*Take, for example, the visual requirements of reading. Observers required a bandwidth of at least 2 cycles per character and a window size of only 4 characters for optimal performance. By the Nyquist theorem, any bandlimited signal may be represented by two samples per cycle of the bandwidth of the signal. Therefore, the display can be represented by $4 \times 16 = 64$ discrete samples.

and resolution. More speculatively, we also supposed that, since vision operates over a large dynamic range, with a commensurate range of signal-to-noise ratios, that the internal representation might take the signal-to-noise ratio into account, using fewer bits (e.g. fewer neural spikes) at lower signal-to-noise ratios. These parameters—size, resolution, and signal-to-noise ratio—can be summarized by a single number, bits, by asking how much information could be transmitted by a visual display with these limitations. So we end up talking about "display" information, in bits, but, like Miller's "message length", it is applicable to a single image, unlike entropy, which can only be defined for a set of possible images.

Bits. Channel capacity is the maximum rate at which a communication channel can transmit information. For continuous signals in the presence of noise, Shannon (Shannon & Weaver, 1949) proved that the capacity C, in bits per sec, of the channel is given by

$$C = W \log_2(1 + P_S/P_N), \tag{1}$$

where W is the bandwidth of the channel, P_N is the power of the white noise perturbing the channel and P_S is the average signal power. If the channel transmits a message only for a duration T, then the information capacity B of the message, in bits, is

$$B = TC. \tag{2}$$

As noted above, any band-limited signal may be represented, without loss of information, as a series of discrete samples taken at twice the bandwidth of the signal (Nyquist, 1928). From equations (1) and (2) we can write the information capacity of the message, B, as the product of the number of discrete samples required to represent the message, $2TW$, and the bits per sample, $0.5 \log_2(1 + P_S/P_N)$. So,

$$B = 2TW\, 0.5 \log_2(1 + P_S/P_N). \tag{3}$$

Consider a display with extents X, Y and T, and bandwidths W_x, W_y and W_t. (X and Y are measured in deg, T in sec, W_x and W_y in c/deg and W_t in c/sec or Hz.) The total number of samples required to represent such a display is $2XW_x 2YW_y 2TW_t$. The total information capacity B of such a visual display with $0.5 \log_2(1 + P_S/P_N)$ bits/sample, is:

$$B = 2XW_x 2YW_y 2TW_t\, 0.5 \log_2(1 + P_S/P_N). \tag{4}$$

Since all the parameters are under direct experimental control, this formula allows the characterization of any set of restrictions by the resulting information capacity B.

When the image is static ($T \to \infty$ and $W_t \to 0$) the required number of temporal samples $2TW_t$ is one, and equation (4) reduces to

$$B = 2XW_x 2YW_y\, 0.5 \log_2(1 + P_S/P_N). \tag{5}$$

Equations (4) and (5) allow the calculation of the information capacity of the display for continuous and static images respectively in terms of the physical parameters of the display.

Display information describes how much information could be sent through a display with specified size, resolution, and noise by an unspecified ensemble of signals with known contrast power. We can compute the "display information" of any image; we only need to know its size, resolution, and signal-to-noise ratio.

Detectability. We note parenthetically that the display information B is closely related to the *energy* signal-to-noise ratio E/N that determines the detectability of a signal known exactly (e.g. Peterson, Birdsall & Fox, 1954; Tanner & Birdsall, 1958; Pelli, 1985). By dividing the signal energy E of a visual image by its spatiotemporal extent we obtain the average signal power P_S. By multiplying the two-sided noise spectral density N by the two-sided spatiotemporal bandwidths of the image we get the average noise power P_N. Therefore,

$$P_S/P_N = \frac{E/XYT}{N 2W_x 2W_y 2W_t} = \frac{E/N}{2XW_x 2YW_y 2TW_t}. \tag{6}$$

The denominator is the number of samples required to represent the image. Thus, the power signal-to-noise ratio P_S/P_N is the energy signal-to-noise ratio E/N per sample. In most of our experiments the power signal-to-noise ratio is small. In this case we can use the identity

$$\lim_{x \to 0} \ln(1 + x) = x$$

as the basis for an approximation, and after substitution, equation (4) becomes

$$B \approx \frac{E/N}{2 \ln 2} \quad \text{if} \quad P_S/P_N \ll 1. \tag{7}$$

This says that, at low power signal-to-noise ratios, the display information in bits is approx. 0.72 times the energy signal-to-noise ratio. This shows a close link between display information and ideal detectability.

Information capacity of the observer

To find the information capacity of a particular attentive task, we minimized display information while retaining near-optimal performance of the task. This involved searching for the optimal combination of spatial extent, temporal duration, spatial and temporal bandwidth and signal-to-noise ratio. For each of the tasks that we considered, we studied the effect of varying each of the relevant display parameters on performance. From these measurements we were able to make an informed choice of the combination of parameter values to use in the measurement of the least display information required for near-optimal performance of the task. The display information is an upper bound on the amount of information used by the observer.

Preliminary studies and the critical-number model

While one can easily detect a single moving dot among hundreds of stationary dots, it is very difficult to detect

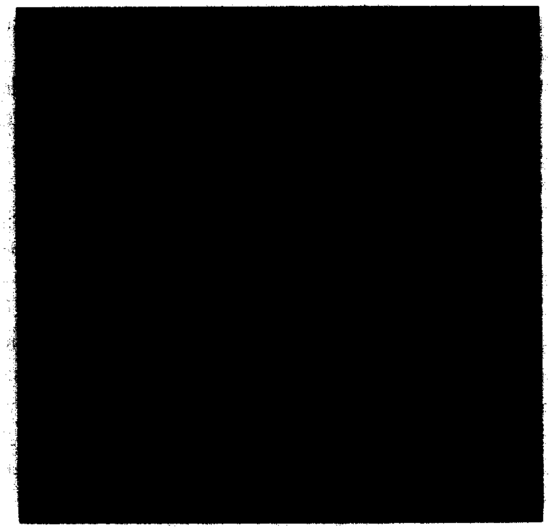

FIGURE 2. A single frame of the dead fly task.

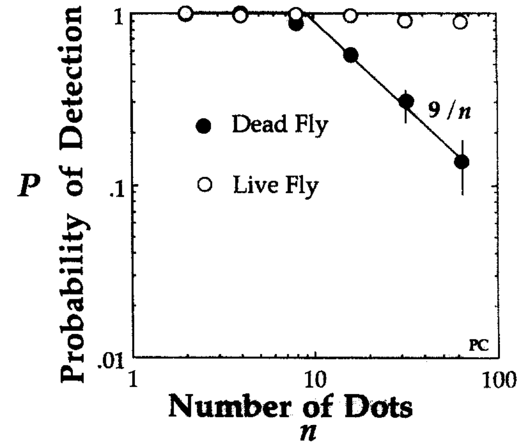

FIGURE 4. Probability of detecting the target fly as a function of the number of dots. The solid symbols are for the dead fly task (attentive) and the open symbols are the live fly task (preattentive). The straight line is the fit of the critical number model to the data for the attentive task. The probability of detection P is 1 when the display has up to the critical number k of dots. When the number of dots n is greater than k then $P = k/n$. The vertical line through each data point is the 95% confidence interval. The observer's initials appear in the bottom right corner of the graph.

a single stationary dot among moving dots (Verghese, Pelli & Barlow, 1988). To measure performance at this task, we display a field of randomly distributed dots (Fig. 2). When the observer is ready, all the dots but one are suddenly displaced in random directions. After the single displacement, the display remains static until the observer responds. The task is to locate the stationary dot. We call it "finding the dead fly". Figure 4 (solid symbols) plots probability of detection vs number of dots in the display, which varied from 2 to 64. When the display contains few dots the observer reliably locates the stationary dot. As the number of dots in the display increases, detection becomes harder and the probability of detection falls.

The data can be conveniently summarized by a one-parameter model; the probability of detection is perfect up to a critical number of dots, k, and then falls as k/n.

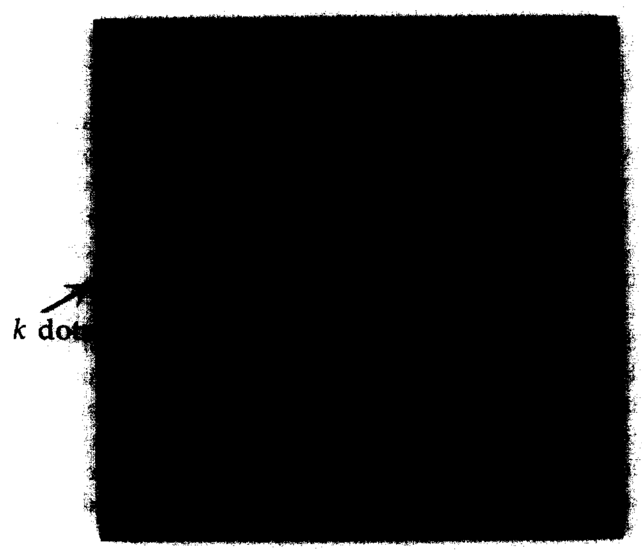

FIGURE 3. The critical-number model for the observer's span of attention. The dashed circle shows the critical number k of dots that an observer can monitor. If there are up to k dots in the display, the observer can monitor the whole display; if there are more than k dots, the observer can monitor only k of them.

This is the *critical-number model*. This model assumes that the observer can monitor up to k dots in the transition from the first frame to the second, as indicated by the dashed circle in Fig. 3. If the target dot is among these, it is successfully detected. The probability of detection is therefore simply the probability that the single target dot is among the k dots that are spanned by the observer's attention. This probability is 1 when $n < k$ (all dots falling within the span of attention) and drops to k/n when $n > k$ (the number of dots exceeding the span of attention). The model is plotted as a solid line in Fig. 4, with a critical number k of 9. The excellent fit to the data suggests that the observer takes in 9 elements in a single span of attention. This model provides a good one-parameter summary of most of our results. Also plotted on the same graph (open symbols) is the probability of detection for the complementary task of finding a live fly. The target is a single moving dot in a field of stationary dots. The task is to locate the single moving dot. In this experiment, the probability of detection is roughly independent of the number of dots (up to 64 dots), which is consistent with motion causing pop out (Nakayama & Silverman, 1986).

(On the first reading we suggest skipping from here to the Discussion section since the flavor of the conclusions is independent of the particular tasks that we studied.)

METHODS

Image generation

All stimuli were generated on an Apple high resolution monochrome monitor driven by a TrueVision video card installed in a Macintosh II computer. The monitor had a resolution of 76 pixels/in. (30 pixels/deg at a viewing distance of 57.5 cm). The video card was used in the 8-bit pixel mode. The luminance of the monitor was measured as a function of the number loaded into the 8-bit digital to analog converters that drove the video card. The measured gamma function describing the non-linear relation between input number and luminance was used to accurately control luminance at the display (Pelli & Zhang, 1991). The background luminance of the display was 25 cd/m^2 and, unless otherwise indicated, observers viewed the display binocularly from a distance of 57.5 cm.

The stimuli (field of dots or checkerboard) were confined to a square window centered on the display. The rest of the screen was at the background luminance. The window size ranged from $0.5 \times 0.5°$ up to $16 \times 16°$.

Temporal duration corresponded to the number of stimulus frames that were shown. The Apple monitor had a frame rate of 67 Hz, so a temporal duration of 1 sec corresponds to a *movie* with 67 frames. Movie durations ranged from 105 msec (7 frames) to 2 sec.

Spatial resolution was controlled by convolving the image with a 2-dimensional circularly symmetric Gaussian. The cut off frequency of the Gaussian filter was taken to be the one-sided bandwidth at which the contrast gain was 0.5. The images were blurred with filter bandwidths ranging from 7.9 down to 0.7 c/deg. The volume (integral over space) of the Gaussian was always one, so its amplitude fell as its bandwidth was increased.

Temporal blurring was achieved by exponentially averaging frames. The contrast of each pixel was replaced by the exponentially weighted average of the contrasts at the same location in previous frames. The exponential weighting function always had unit integral. This is similar to the effect of a slow phosphor, but scaled to preserve light. The time constant of the exponential decay depended on the desired degree of temporal blur. The cut off frequency of the exponentially decaying filter was taken to be the one-sided bandwidth at half peak amplitude. Temporal bandwidths ranged from 18.4 to 3.7 Hz.

Signal-to-noise ratio was varied by adding white noise to the display. "White" noise refers to the flat shape of the power spectrum (by a loose analogy to the spectrum of white light). Specifically, the luminance of each pixel was perturbed by adding a zero-mean random number drawn from a uniform distribution. The maximum signal and noise amplitudes were each constrained to half of the available gray scale.

We follow the modern convention of describing visual stimuli such as the signal $s(x, y, t)$ and noise $n(x, y, t)$ as contrast functions, that is, as deviations from the background luminance L_b, normalized by the background luminance. Thus, the luminance function $L(x, y, t)$ over space and time of the display is:

$$L(x, y, t) = [1 + s(x, y, t) + n(x, y, t)]L_b. \quad (8)$$

The visual noise $n(x, y, t)$ is random fluctuations in luminance over space and time, introduced by the experimenter. The contrast power of the signal and noise are their average squared values:

$$P_S = X^{-1}Y^{-1}T^{-1}\int_{-X/2}^{x/2}\int_{-Y/2}^{y/2}\int_{-T/2}^{t/2} s^2(x, y, t)\,dx\,dy\,dt, \quad (9)$$

$$P_N = X^{-1}Y^{-1}T^{-1}\int_{-X/2}^{x/2}\int_{-Y/2}^{y/2}\int_{-T/2}^{t/2} n^2(x, y, t)\,dx\,dy\,dt, \quad (10)$$

where XY is the area of the smallest rectangle bounding the signal, and T is the duration of the signal presentation. The average power is the mean square contrast c_{rms}^2, which is often used to describe visual noise (e.g. Stromeyer & Julesz, 1972; Pelli, 1981, 1990).

Signal-to-noise ratio P_S/P_N, is the ratio of average signal power to average noise power. Signal-to-noise ratios were calculated for the area-duration of the signal, XYT. In the dot experiments XY, the smallest rectangle bounding the signal, was the area of the display window. In the check experiments XY was the smallest rectangle enclosing the checks in the display. The dependence of signal-to-noise ratio on the number of elements was different for dots and checks as we kept the display window fixed for dots, and the check size fixed for checks. In the dot experiments, signal power and hence signal-to-noise ratio were proportional to the number of dots. In the check experiments, the signal and noise powers, and hence the signal-to-noise ratio were independent of the number of checks. Signal-to-noise ratios ranged from 0.00018 to infinity (i.e. no noise).

We pre-computed all the stimuli and stored them on the computer's disk (i.e. slow storage). Before each trial we loaded all the required images into computer RAM (i.e. fast storage). The computer moved data from RAM to the video display card quickly enough to show movies with 256×256 pixels (with a pixel depth of 8 bits/pixel) in real time, i.e. a new image on every frame, at 67 Hz.

In the dot experiments small dots were randomly distributed over the display window. We were concerned that the luminance of horizontally adjacent dots might be artifactually high as a result of the monitor's finite video slew rate (Lyons & Farrell, 1989; Pelli & Zhang, 1991). The beam can brighten and darken only slowly, taking more time to swing full scale than it takes to sweep across a single pixel. Since the slew rate limitation precedes the nonlinear gamma, this might result in a net luminance error. We checked to see if the slew rate caused horizontally adjacent dots (2×2 pixel squares) to be more than twice as intense as a single dot. To do this, we compared the luminances of two vertical gratings, one with alternating 2-pixel bright and dark bars (simulating isolated dots) and one with alternating 4-pixel bright and dark bars (simulating adjacent dots). (The dark and light bars were at the same luminances, respectively, as the background and the dots used in the experiments.) Both gratings had the same mean luminance, indicating that the monitor's finite slew rate did not introduce any luminance artifacts into the dot experiments.

Experimental paradigms

We measured information capacity for two kinds of attentive task: detection of a static dot in a field of moving dots and detection of a non-flashing square in a field of flashing squares. For both tasks we studied the effects of spatial extent, spatial bandwidth (degree of blur), signal-to-noise ratio, and, where appropriate, duration and temporal bandwidth. For each experimental condition we measured probability of detection as a function of the number of distractors. Each such graph can usually be summarized by a critical number k.

Detecting the absence of motion: "finding the dead fly". The first frame had a variable number of dots (2–64) of fixed size located randomly in the display window, as in Fig. 2. In subsequent frames, the target dot remained stationary, while each of the other dots was displaced by a fixed amount (4 pixels or 8′ of arc at a viewing distance of 57.5 cm) from frame to frame. Each dot's direction of displacement was random. We didn't specifically prevent overlap between dots, as doing so would have introduced dependencies between dots. When dots overlapped they obscured one another, as reflective white squares would, rather than summing as light sources would. The observer's task was to locate the target dot. We measured the probability of detection for two versions of this task: the single displacement "two-frame" case and the continuous "many-frame" case. In the *two-frame* case, observers viewed the first frame for as long as they wanted and initiated display of the second frame by hitting a key. Then the first frame was replaced by the second and the observer was then asked to use the cursor to locate the target dot. The cursor was displayed as a white arrow superimposed upon the static field. The observer moved the cursor by manually sliding a "mouse". The dot closest to the cursor when the observer clicked the mouse button was taken to be the dot chosen by the observer. The observer was given auditory feedback indicating a right or wrong response and visual feedback indicating the location of the target dot. All the information about the movement of the dots was contained in the transition from the first frame to the second frame. For the *many-frame case*, the observer viewed a movie of fixed duration. The observer initiated the movie, and at the end of the movie was asked to locate the target dot in the last frame, which remained frozen on the display until the observer responded.

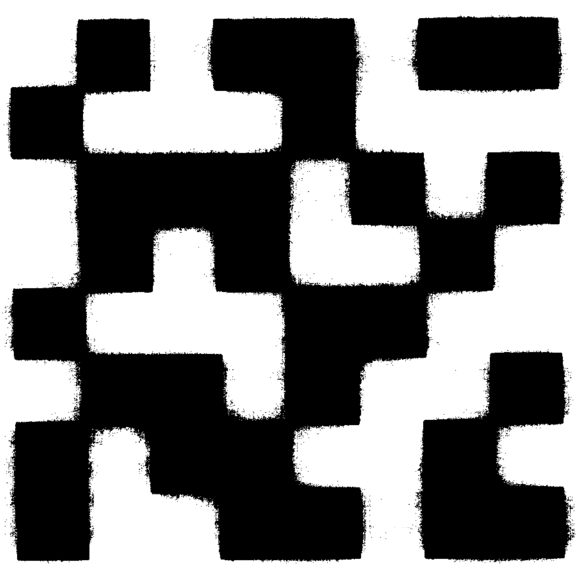

FIGURE 5. A single frame of the dead firefly task.

We also repeated the experiments with a larger dot displacement of 20′ of arc between frames. The data for the two-frame case show that the detection of motion remains preattentive and the absence of motion attentive, irrespective of the displacement between frames, at least for these two sizes is displacement, 8′ and 20′ (Verghese *et al.*, 1988). The rest of the experiments used a dot displacement of 8′.

Detecting the absence of change: "finding the dead firefly". The stimuli consisted of a random checkerboard (Fig. 5). From frame to frame, each check randomly changed polarity from dark to light or vice versa, except for the target check, which remained static. The observer's task was to locate the target check. The check that contained the tip of the cursor when the mouse was clicked was taken to be the observer's choice. Both the two-frame and many-frame versions of the task were examined. In the two-frame case, however, all checks except the target changed polarity in the second frame to avoid the possibility of more than one static check.

Initially we kept the overall size of the window constant, decreasing the size of the individual checks as their number increased. However, we found that location accuracy fell with number of elements much more slowly than predicted by the critical number model. Upon consideration we realized that as the number of checks increased in a display window of fixed size, the spatial frequency bandwidth of the display increased. The higher spatial frequency content of the small checks might improve performance in the many-frame case by increasing the visual integration time (Barlow, 1958; Robson, 1966). This interpretation was supported by finding that keeping check size constant (varying window size with the number of checks) abolished the discrepancy, yielding data well fit by the critical number model. Therefore, we used a constant check size in the rest of the experiments reported here.

Preattentive tasks. We also measured probability of detection for preattentive versions of these two tasks, i.e. for the detection of motion and the detection of change. The stimulus presentation was modified so that the target was the dot that moved in a field of stationary dots, or a check that flashed in an otherwise static checkerboard. In terms of creating the display, this meant that the first frame was similar to that for the attentive experiments, but that in subsequent frames only the target dot moved or the target check flashed, while all the other elements in the display remained static.

Unsuitability of the many-frame experiments. In the many-frame versions of the dead fly and dead firefly experiments, temporal resolution and duration were new variables, in addition to the parameters in the two-frame case. These temporal blur results reveal that our many-frame tasks are inappropriate for the measurement of the least information capacity required. (For complete results of these experiments see Verghese, 1990.) Temporal averaging improves performance in the many-frame version of each of these tasks. As temporal blur increases, the dynamic distractors become fainter and fainter while the static target is unaffected. The contrast difference between the static target and the faint dynamic distractors increases with temporal blur, until the static target "pops out" at large blurs. Increasing temporal blur would in the limit eliminate the dynamic distractors altogether, leaving just the static target. The minimum information capacity would require the maximum temporal blur and this eliminates the temporal dimension altogether, so nothing at all was gained by introducing the temporal dimension to these tasks.

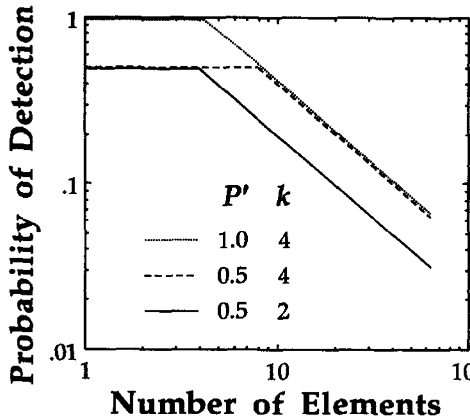

FIGURE 6. The predicted probability of detection P_{detect} vs the number of elements n is plotted for the two-parameter version of the model [equation (15)]. The two parameters are the critical number k and the maximum probability of detection P'.

Probability of detection and the critical-number model

For every task we measured percent correct location of the target as a function of the total number of elements in the display. In our model, the measured probability $P_{correct}$ is the sum of the true probability that the target was detected, and the probability of guessing the target (by chance) when it was not detected

$$P_{correct} = P_{detect} + (1 - P_{detect})P_{guess}, \tag{11}$$

which we can solve for P_{detect}

$$P_{detect} = \frac{P_{correct} - P_{guess}}{1 - P_{guess}}. \tag{12}$$

The measured proportion of correct responses was corrected for guessing by equation (12) in order to estimate the probability of detection P_{detect} that is plotted in the figures. The probability of guessing the target correctly, P_{guess}, was assumed to be the reciprocal of the number of elements in the display,

$$P_{guess} = \frac{1}{n}. \tag{13}$$

These data were then fit (by eye) by the critical number model, adjusting k for best fit.

$$P_{detect} = \begin{cases} 1\ n & \text{if } n \leqslant k \\ k/n & \text{if } n > k \end{cases}. \tag{14}$$

A two-parameter version of the critical-number model

In a few conditions this one-parameter critical number model did not fit the data. This model predicts perfect detection up to the critical number k. However, for some conditions (small signal-to-noise ratios and high degrees of spatial blur) probability of detection was well below 100% even for a single element. If the observer monitored k' elements, but the probability of detecting the target among those elements, P', was <1, then the observer would detect the target with probability P' if there were less than k' elements and would detect the target with probability $P'k'/n$ if there were more than k' elements. For this model we found it convenient to define $k = P'k'$, so that the probability of detection is

$$P_{detect} = \begin{cases} P' & \text{if } n \leqslant k/P' \\ k/n & \text{if } n > k/P' \end{cases}. \tag{15}$$

The behavior of this model is plotted in Fig. 6. P' controls the height of the horizontal section of the curve. The dotted line has a P' of 1; the dashed line has a P' of 0.5. As in the one-parameter model, k is uniquely determined by the falling section of the curve. The broken lines have a k of 4 and the solid line has a k of 2.

Observers

A total of five observers participated in this study: four in the preliminary experiments and two in the measurement of the information capacity of visual attention. The observers were students, 20–28 yr of age, with normal visual acuity or acuity corrected to normal. Three of the observers (PR, PC, WP) were unaware of the purpose of the experiments. We have a complete set of data for two observers, but will show the effect of varying display parameters for only one. Data for both observers are shown for the combination of parameters that give the least amount of display information required for near optimal performance.

RESULTS

We will adopt the following format to describe our results from the four experiments. In each experiment, we first studied the effect of varying each of the display parameters. For each condition we measured the probability of detection as a function of the number of elements (target plus distractors) and fit the data with the critical-number model. For most conditions the one-parameter version of this model is a good fit; in a few conditions the data require the two-parameter version. In the interests of brevity we will not plot the raw data for each of these conditions nor the critical number fits (for details, see Verghese, 1990). Instead, we will summarize this vast amount of data by reporting the parameter values that gave the best* critical number. Each experiment suggested a combination of restrictions that would minimize the display information while leaving performance almost unimpaired as compared with the unrestricted condition. The final graphs of this section directly compare the unrestricted performance with that of the combined restrictions.

Experiment 1. "Finding the dead fly". Detecting the absence of motion: two-frame case

Observers were presented with a two-frame display of dots. In the transition from the first (static) frame to the second, all the dots but one were displaced in random directions. The observer's task was to locate the single static dot.

Effect of window size, spatial blur and signal-to-noise ratio. We varied window size by either changing the display or the viewing distance. For both variations, the critical number increases with decreasing window size, peaks at a window size of $4.3 \times 4.3°$, and then falls. We set the display window at $4.3 \times 4.3°$ for all subsequent conditions of this experiment, and for the measurement of the information capacity. Blur was achieved by low pass filtering the display (Gaussian convolution). Critical number improves with increasing blur, up to a spatial bandwidth of 4 c/deg, and then falls. This is the value of blur we used in the combined restriction condition. Signal-to-noise ratio was varied by adding zero-mean noise to the display. Critical number increases with increasing signal-to-noise ratios, and then levels off. As this function does not have a peak, we arbitrarily chose a signal-to-noise ratio of $0.012n$ for the combined restriction condition.

Effect of combined restrictions. Figure 7 shows examples of the unrestricted and the restricted stimulus in the two-frame dot task. The solid symbols in Fig. 8(a) show probability of detection for observer PV for the combination of the following display parameters: a display window of $4.3 \times 4.3°$, a spatial bandwidth of 4 c/deg and a signal-to-noise ratio of $0.012n$. The open symbols on the same graph show the probability of detection in the unrestricted condition. The combined restrictions reduce the critical number from 9 to 6, only a small reduction. The display information under the combined restrictions when the critical number of dots is displayed, can be calculated from equation (5) and is 58 bits. The signal-to-noise ratio is $0.012n = 0.012 \times 6$ since we consider the condition where $n = k$, which is 6 in this experiment.

We set $n = k$, that is, set the number of elements to the critical number, because our general approach in measuring information capacity is to restrict each dimension (including number of elements) as much as possible without impairing performance. Observer LZ shows chance performance at this signal-to-noise ratio ($0.012n$). The display parameter values required for near optimal performance in her case were a signal-to-noise ratio of $0.048n$ and a spatial bandwidth of 2.6 c/deg [Fig. 8(b)]. Information capacity for this combined restriction condition is 57 bits. Although the critical number of dots is different for the two observers, the information at the display required for near optimal performance of this task is nearly identical.

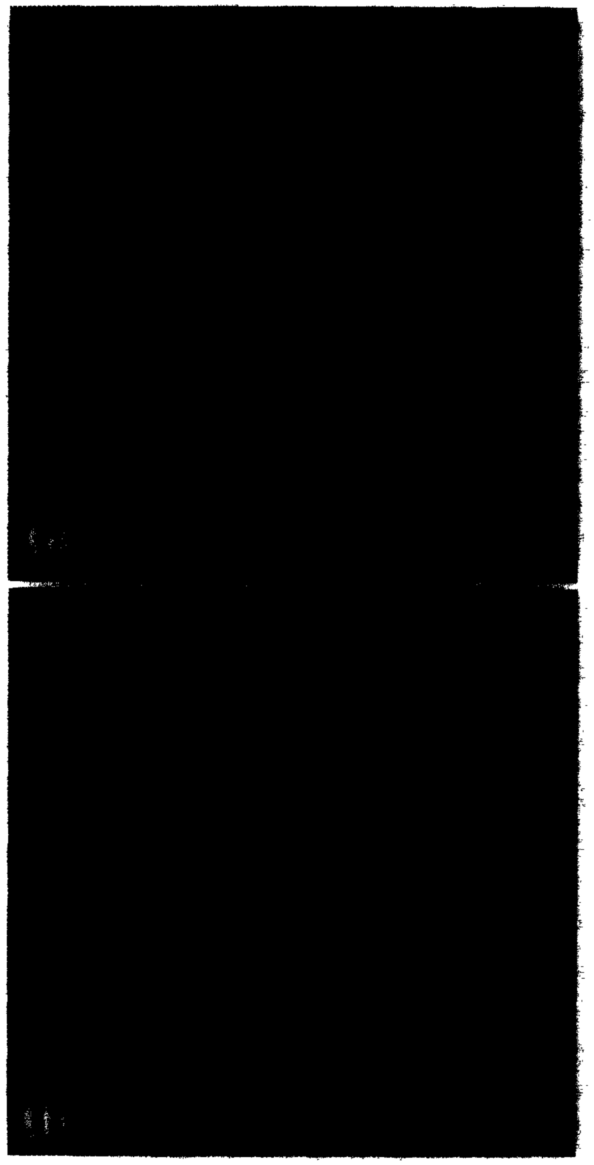

FIGURE 7. Examples of stimuli used in the unrestricted condition (a) and the restricted condition (b) of the dot experiments. In (b) the spatial filter cut-off is 4 c/deg and the signal-to-noise ratio is $0.003n$. At a viewing distance of 67 cm (2′ 3″), these figures subtend $4.3 \times 4.3°$, the display size used in most of the dot experiments.

Experiment 2. "Finding the dead firefly". Detecting the absence of change: two-frame case

Observers were presented with a two-frame display of checks. In the transition from the first frame to the second, all the distractor checks changed polarity from light to dark or vice versa. The target check was left unchanged. The observer's task was to locate the target, the single static check. In the two-frame case, the probability of detection is high even at a large number

*"Best" typically refers to the peak of the function describing critical number versus a given display parameter. For some display parameters, this function has no clear peak. In such cases, we chose the lowest parameter value that gave unimpaired performance.

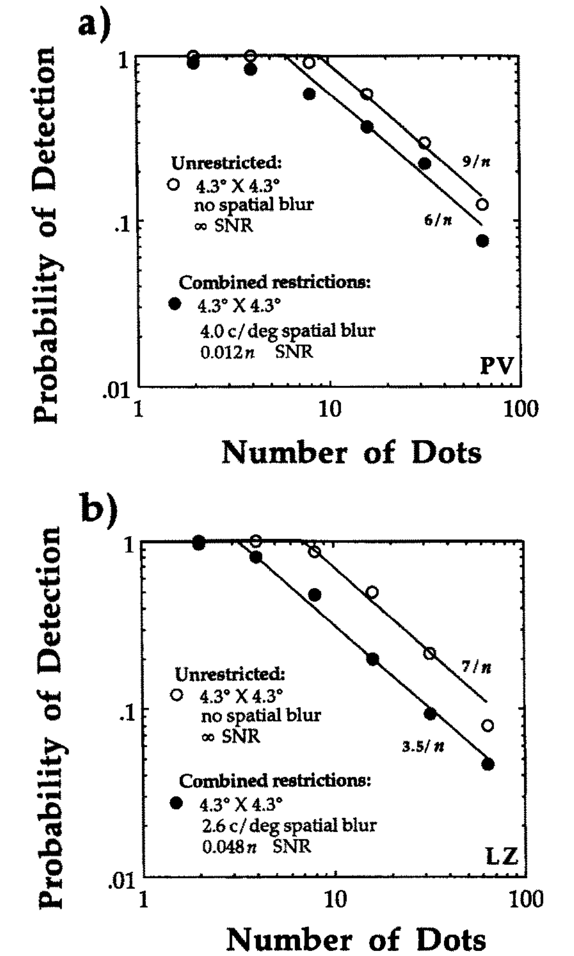

FIGURE 8. Probability of detection vs the number of dots for the unrestricted (open symbols) and combined restriction (solid symbols) conditions of the two-frame dot experiment for observers PV and LZ. SNR is the signal-to-noise ratio.

of checks. Observers report that they memorize the pattern in the first frame. In the second frame, when all the checks change polarity except the target check, the observers report that they look for the pattern they memorized to change polarity. The part of the pattern that does not look like the contrast reversed version of the pattern they saw in the first frame is where they suppose the target check is.

Effect of check size, spatial blur and signal-to-noise ratio. In the check tasks, probability of detection was measured as a function of check size, rather than window size *per se*. (Window size depends on the size of the checks and the number of checks in the display.) Critical number improves slightly with increasing check size, up to a check size of 0.26 × 0.26° and then drops. We used this check size for the rest of Experiment 2. With spatial blur, the critical number initially improves and then deteriorates with increasing blur. The critical number is highest at a spatial bandwidth of 2.75 c/deg, so we used that value in the combined restriction condition. The signal power for the check tasks is taken as the average power over the area of the checks in the display. Thus the signal-to-noise ratio remains constant as the number of checks increases. Critical number decreases with the addition of noise. A signal-to-noise

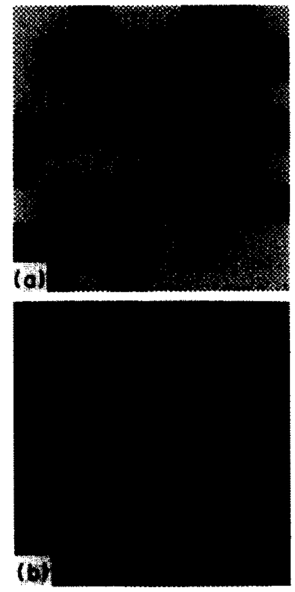

FIGURE 9. Examples of stimuli used in the unrestricted condition (a) and the restricted condition (b) of the check experiments. In (b) the spatial filter cut-off is 4 c/deg and the signal-to-noise ratio is 0.86. At a viewing distance of 67 cm (2′ 3″), these figures subtend 2.1 × 2.1°, the display size used in most of the check experiments.

ratio of 1.52 is the lowest value that gives almost unimpaired performance. This is the value that we used in the combined restriction condition.

Effect of combined restrictions. Figure 9 shows examples of the unrestricted and restricted stimulus in the two-frame check task. Figures 10(a) and (b) plot probability of detection for the unrestricted condition and the combined restriction condition, for observers PV and LZ respectively. The combination of restrictions for both observers is: a check size of 0.26 × 0.26°, signal-to-noise ratio of 1.52 and filter bandwidth of 2.75 c/deg. For observer PV critical numbers are 27 and 29 for the restricted and unrestricted conditions respectively. The information capacity of the display in the restricted condition is 39 bits, yet yields almost unimpaired performance. Figure 10(b) is LZ's performance for the same condition. LZ attained the same critical number (20) in the combined restriction condition as in the unrestricted condition. The information capacity in the combined restriction condition is 29 bits.

Experiment 3. Information capacity at different spatial scales

To study the effect of spatial scale on the least information required for an attentive task, we chose the two-frame version of finding the dead firefly (Experiment 2). We redid the experiment at two other scales—with check sizes of 0.066 and 1.04°. Figure 11(a) and (b) plots the results for these two scales. As in the previous graphs, the open symbols plot probability of detection for the unrestricted condition and the solid symbols for the combination of restrictions that allow almost unimpaired performance. The display information required for the smaller scale (check size of 0.07°) is 38 bits and for the larger scale (check size of 1.04°) is 27 bits.

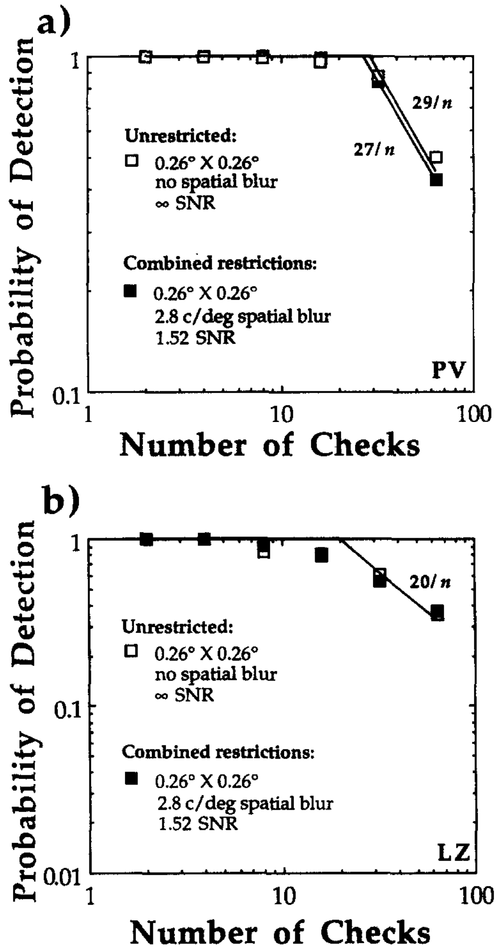

FIGURE 10. Probability of detection vs the number of checks for the unrestricted (open symbols) and restricted (solid symbols) conditions of the two-frame check experiment for observers PV and LZ. SNR is the signal-to-noise ratio.

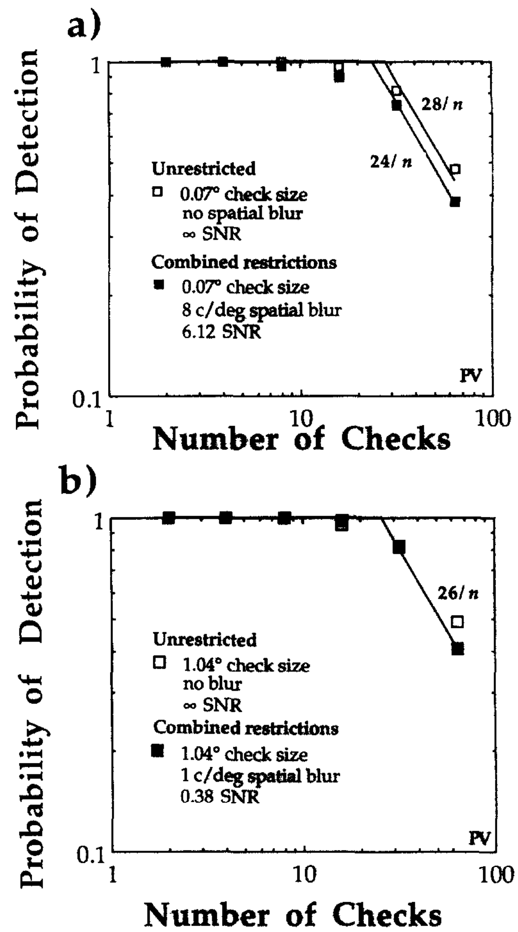

FIGURE 11. Probability of detection vs number of checks for two check sizes: (a) 0.07° and (b) 1.04°. The open symbols are for the unrestricted condition and the solid symbols are for the restricted condition. SNR is the signal-to-noise ratio.

Figure 12 plots display information in bits vs check size. (The display information required for a check size of 0.26° is 39 bits; this was measured in Experiment 2.) The function is relatively flat indicating that the information capacity of visual attention is only weakly dependent on scale.

Experiment 4. Information required for a preattentive task

Here we measured probability of detection for the two-frame case of detecting change: "finding the live firefly". The task is to find the single flashing check in an otherwise static checkerboard. Our preliminary experiments (Verghese *et al.*, 1988) suggest that this is a preattentive task in which the probability of detection is roughly independent of the number of distractors. This is true for up to 64 distractors (Figs 4 and 13). As the number of distractors is increased beyond 64, the probability of detection begins to fall (Fig. 13). However, the probability of detection falls off more slowly than predicted by the critical number model in both the unrestricted and restricted conditions. The critical number model (solid line in Fig. 13) is a poor fit to these data, so we use it only as a rough index to calculate the display information required. (The critical number that best fits the data is 200.) The least display information required for the preattentive task is 2106 bits. This is about two orders of magnitude larger than the display information required for the attentive tasks that we studied.

DISCUSSION

Analysis of results

The critical-number model. Most of our data are well fit by a single parameter, the critical number k.

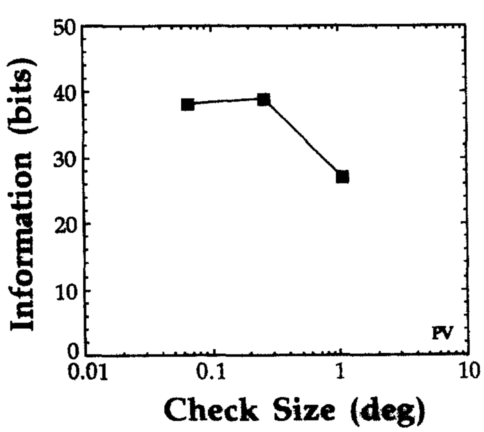

FIGURE 12. Required display information vs check size. Based on the data of Figs 10 and 11.

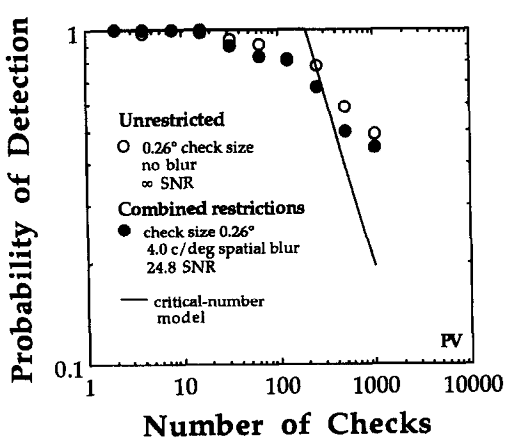

FIGURE 13. Information required for a preattentive task. Probability of detection is plotted vs number of checks for the two-frame case of detecting the flashing square (the live firefly task). Data for the unrestricted condition are shown in open symbols and for the restricted condition in solid symbols. These data are not well fit by the critical number model.

Of the 70 data sets in Experiments 1 and 2, only one could not be well fit, three required the two-parameter version for a good fit, and the rest were well fit by the one-parameter version of the critical-number model. Estimates of k ranged from 0.3 to 39, while estimates of P' ranged from 0.13 to 1.0, over the entire set of data collected in these experiments and across both observers. By assuming stochastic independence of successive trials we estimated the 95% confidence intervals of the parameter values of the critical number model for each data set. These intervals ranged from 0.3 ± 0.2 to 39 ± 4 for k and 0.13 ± 0.04 to 0.99 ± 0.01 for P'. In general, the critical number model offers a parsimonious and useful summary of search accuracy.

Information required for an attentive task. Table 1 shows that for both observers the least display information required for near perfect performance of both two-frame attentive tasks is 44 ± 15 bits. This variation, from 30 to 60 bits, is nontrivial, but seems small when comparing this attentive information capacity of roughly 44 bits with Experiment 6's estimated preattentive information capacity of 2106 bits.

Effect of scale. The least information required at three different scales, spanning a 12 to 1 range, is roughly the same (Fig. 12), ranging from 27 to 39 bits. This result is a quantitative confirmation of the spotlight metaphor that has been used in attention (Norman, 1968; Posner, Synyder & Davidson, 1980; Kahneman & Henik, 1981; LaBerge, 1983). Attention, like a spotlight, can be zoomed in or out, but the power of the beam, like the information capacity of attention, is independent of scale.

TABLE 1. Least display information required

	Dots	Checks	
PV	58	39	bits
LZ	57	29	bits

Information required for a preattentive task. The probability of detecting the live firefly is not independent of the number of distractors when the number of distractors is large (over 200). But it is better than predicted by the critical number model that is a good fit to the data from our attentive tasks. The probability of detection in this task appears to lie between attentive (described by a critical number) and preattentive (independent of number). This and our temporal blur results suggest that there may a continuum rather than a clear dichotomy between attentive and preattentive tasks. Bergen and Julesz (1983), Duncan and Humphreys (1989), and Verghese and Nakayama (1991) have also suggested such a continuum.

Relation to previous work

Preattentive and attentive processing. Both Treisman's feature integration theory and Julesz's texton theories suggest task-dependent descriptions of visual attention. They assume that, in the absence of features or texton gradients that allow the target to be discriminated preattentively, serial search involving visual attention is required. Performance at these search tasks has been measured as the time to detect the target as a function of the number of elements in the display. The slope of this line gives processing time per element, and is an index of the difficulty of the attentive task. Typical search times are of the order of 30–60 msec per element (Bergen & Julesz, 1983). The search times can be as low as 10 msec per item when the stimulus elements are members of an overlearned set such as letters of the alphabet (Sperling, Budiansky, Spivak & Johnston, 1971). In the tasks we considered, search times to detect the static target were 14 msec per dot and 19.5 msec per check in the continuous (many-frame) versions of these tasks. Similarly, critical number is inversely related to difficulty. Both reaction time and critical number are task-dependent descriptions of attention. We sought a task-independent measure of the limited processing capacity of the attentive mechanism.

Our working hypothesis is that visual attention can be characterized by an information capacity that is a fixed, modest number of bits. Our experiments support this hypothesis and indicate that this capacity is at most 44 ± 15 bits. Our experimental results yield the same modest capacity for the attentive processor in two tasks of very different difficulty. For the easy task (detecting the static check), the number of elements that can be processed in a single attentive span is large ($k = 29$). For the hard task (detecting the stationary dot), this number is small ($k = 6$), yet roughly the same amount of display information (bits) is required for both tasks.

Absolute judgment tasks. Measures of the information capacity of the observer range from 2 to 3 bits for unidimensional stimuli and up to 7.2 bits for multidimensional stimuli (Miller, 1956). In our experiments, the observer transmits $\log_2 k$ bits (i.e. which of the k elements is the target), which is 1–6 bits depending on k. We think this measures a late bottleneck—the observer's ability to categorize stimuli—whereas the display-information approach measures what the observer can take in. Indeed, our estimate of 44 ± 15 bits is consistent with Sperling's (1960) estimate of 40 bits for the iconic store, which he measured by his partial-report technique.

Visual requirements of everyday tasks. The hypothesis advanced by this paper is that all attentive visual tasks are limited by the same information capacity. The earlier work on reading by Legge *et al.* (1985) and on mobility by Pelli (1986) indicate that a modest number of samples suffice for these important tasks, but we do not know what signal-to-noise ratios would be appropriate to compute their information capacities.

Other dimensions

It might be argued that the modest display information required by our tasks is due to the fact that our experiments do not allow variations in other visual dimensions—such as color, orientation or spatial scale. To address this issue, Verghese and Pelli (1992) studied the effect of mixing two different spatial scales. Observers were presented with a variation of the dead firefly experiment; the window was divided into two halves that were at different spatial scales. Probability of locating the dead firefly was measured for different ratios of spatial scale. Performance, on average, decreases by a factor of 0.6 when the observer simultaneously monitors scales that are a factor of two apart. This is similar to a result by Sperling and Melchner (1978) for locating numbers among letters.

Would our estimate of the least display information required increase significantly if we included color as a dimension in our attentive tasks? One way to address this question is to ask how much extra information is required to convert a monochromatic display to color. Consider broadcast television—it is made up of one luminance signal and two chromatic signals (red–green and blue–yellow). The NTSC system has a luminance signal bandwidth of 4.2 MHz and a total chromatic bandwidth of 1.7 MHz, which is only about 40% of the luminance bandwidth. However, even 40% is too much. NTSC does not take full advantage of the following facts about color: human spatial and temporal sensitivity to color signals is lower; humans sensitivity to quantization errors in color is lower; and scene statistics lead to less information being physically present along the chromatic axes. Research into digital transmission of color signals and image coding techniques suggest that only about 10–20% more bits are required to code a color image to satisfy human viewers. For instance, with transform coding the color components of the NTSC signal can be transmitted at 14% of the bit rate used for the luminance signal (Netravali & Haskell, 1988). As color requires only a small fraction of the bandwidth required to represent luminance signals, we expect that adding color to an attentive task will not significantly increase the display information required.

CONCLUSIONS

The critical number model uses the idea of a span of attention to provide a quantitative account for the observer's probability of detecting the target in an attentive task. The critical number k is the number of elements that an observer acquires in a glimpse. This complements the more traditional processing time per element used to characterize reaction-time experiments. Nearly all our data for attentive tasks are well fit by the critical number model. The critical number decreases (and the slope of the reaction time function increases) with the degree of difficulty of the task. A two-parameter version of the critical number model—allowing for unreliable detection of the monitored elements—allowed us to fit *all* our results, even when the stimuli were difficult to resolve.

We conjectured that all attentive visual tasks are performed by a single general-purpose processor with a low data capacity. We used Shannon's theorem [equation (4)] to measure the information capacity of the attentive visual system in terms of the physical parameters of the display, namely spatial extent, spatial resolution, and signal-to-noise ratio. For two attentive visual tasks, one easy and one hard, we found the information capacity of attention to be 44 ± 15 bits. This is small when compared to our estimate of 2106 bits for a preattentive task. Our experiments indicate that the attentive information capacity is only weakly dependent on the scale of the display, over a 16:1 range. The conjecture remains unproven, but invites further testing. It would be rejected by evidence that there exists any attentive task that requires substantially more than 44 bits of display information.

REFERENCES

Atkinson, J., Campbell, F. W. & Francis, M. R. (1976). The magic number 4 ± 0: A new look at visual numerosity judgements. *Perception, 5,* 327–334.

Atkinson, J., Francis, M. R. & Campbell, F. W. (1976). The dependence of the visual numerosity limit on orientation, colour, and grouping in the stimulus. *Perception, 5,* 335–342.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology, 141,* 337–350.

Bergen, J. R. & Julesz, B. (1983). Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man and Cybernetics, 13,* 857–863.

Crossman, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology, 5,* 41–51.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Fiorentini, A., Maffei, L. & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12,* 195–201.

Ginsberg, A. P. (1980). Specifying relevant spatial information for image evaluation and display design: An explanation of how we see certain objects. *Proceedings SID 21,* 219–227.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11–26.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3,* 281–282.

Julesz, B. (1981a). A theory of preattentive texture discriminations based on the first-order statistics of textons. *Biological Cybernetics, 41,* 131–138.

Julesz, B. (1981b). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97.

Julesz, B. (1984). Towards an axiomatic theory of preattentive vision. In Edelman, G. M., Gall, W. E. & Cowan, W. M. (Eds) *Dynamic aspects of neocortical function* (pp. 585–612). New York: Neurosciences Research Foundation.

Kahneman, D. & Henik, A. (1981). Perceptual organization and attention. In Kubovy, M. & Pomerantz, J. R. (Eds), *Perceptual organization* (pp. 181–211). Hillsdale, N. J.: Earlbaum.

Kaufman, E. L., Lord, M. W., Reese, T. W. & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 498–525.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 371–379.

Legge, G. E., Rubin, G. S. & Luebker, A. (1987). Psychophysics of reading. V. The role of contrast in normal vision.

Lyons, N. P. & Farrell, J. E. (1989). Linear systems analysis of CRT displays. *SID Digest, 20,* 220–223.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Miller, G. A. (1957). Human memory and the storage of information. *IRE Transactions on Information Theory, IT-2,* 129–137.

Nakayama, K. & Silverman, G. H. (1986). Serial and parellel processing of visual feature conjunctions. *Nature, 320,* 264–265.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Netravali, A. & Haskell, B. (1988). *Digital pictures* (p. 438). New York: Plenum Press.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522–536.

Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the A.I.E.E., 47,* 617–644.

Owsley, C., Sekuler, R. & Boldt, C. (1981). Aging and low-contrast vision: Face perception. *Investigative Ophthalmology and Visual Sciences, 21,* 362–365.

Pavel, M., Sperling, G. Riedl, T. & Vanderbeek, A. (1987). Limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America A, 4,* 2355–2365.

Pearson, D. E. (1983). Evaluation of feature-extracted images for deaf communication. *Electronics Letters, 19,* 629–631.

Pearson, D. E. & Robinson, J. A. (1985). Visual communication at very low data rates. *Proceedings of the IEEE, 73,* 795–812.

Pelli, D. G. (1981). *Effects of visual noise.* Doctoral dissertation, Cambridge University, Cambridge, England.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America, 2,* 1508–1532.

Pelli, D. G. (1986). The visual requirements of mobility. In Woo, G. C. (Ed.), *Low vision: Principles and applications. Proceedings of the international symposium on low vision* (pp. 134–145). Springer: University of Waterloo.

Pelli, D. G. (1990). The quantum efficiency of vision. In Blakemore, C. (Ed.), *Visual coding and efficiency* (pp. 3–24). Cambridge: Cambridge University Press.

Pelli, D. G. & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31,* 1337–1350.

Pelli, D. G., Legge, G. E. & Schleske, M. M. (1985). Psychophysics of reading III. A fiberscope low-vision reading aid. *Investigative Ophthalmology and Visual Sciences, 26,* 751–763.

Pelli E., Goldstein, R. B., Trempe, C. L. & Arend, L. E. (1989). Image enhancement improves face recognition. In *Topical meeting on noninvasive assessment of the visual system, digest series* (Vol. 7, pp. 64–67). Washington, D.C.: Optical Society of America.

Peterson, W. W., Birdsall, T. G. & Fox, W. C. (1954). Theory of signal detectability. *Transactions of the IRE PGIT, 4,* 171–212.

Posner, M. I., Snyder, C. R. R. & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America, 56,* 1141–1142.

Schuchard, R. A. & Rubin, G. S. (1989). Face identification of low pass filtered faces and letters. *Investigative Ophthalmology and Visual Sciences, 30 (Suppl.),* 396.

Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication.* University of Illinois.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74(11),* 1–29.

Sperling, G. & Melchner, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science, 202,* 315–318.

Sperling, G., Budiansky, J., Spivak, J. G. & Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for a numeral. *Science, 174,* 307–311.

Sperling, G., Landy, M. S., Cohen, Y. & Pavel, M. (1985). Intelligible coding of ASL images at extremely low information rates. *Computer Vision, Graphics and Image Processing, 31,* 335–391.

Stromeyer, C. F. III & Julesz, B. (1972). Spatial frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America, 62,* 1221–1232.

Tanner, W. P. Jr & Birdsall, T. G. (1958). Definitions of d' and η as psychophysical measures. *Journal of the Acoustical Society of America, 30,* 922–928.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing, 31,* 156–177.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Verghese, P. (1990). *The information capacity of visual attention* (Report No. ISR-S-27). Syracuse, N.Y.: Institute for Sensory Research, Syracuse University.

Verghese, P. & Nakayama, K. (1991). Implications of a limited processing capacity for visual search tasks. *Investigative Ophthalmolgy and Visual Sciences, 32 (Suppl.),* 1041.

Verghese, P. & Pelli, D. G. (1990). The information capacity of visual attention. *Investigative Ophthalmology and Visual Sciences, 31 (Suppl.),* 562.

Verghese, P. & Pelli, D. G. (1992). The spatial frequency bandwidth of attention. *Vision Research.* Submitted.

Verghese, P., Pelli, D. G. & Barlow, H. B. (1988). Detecting the absence of motion. *Investigative Ophthalmology and Visual Sciences, 29 (Suppl.),* 406.

Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Acknowledgements—We thank Bart Farrell, Mary Hayhoe and Ken Nakayama, for helpful comments. The discussion of other dimensions reflects helpful discussions with Ted Adelson and Oliver Braddick. This work was supported by NIH grant EY04432 and NASA Ames University Consortium Joint Research Agreement NCA2-232 to Denis Pelli, and by an Allyn Foundation Inc. grant to the Institute for Sensory Research. Preeti Verghese was supported by grant AFOSR-90-0330 to Ken Nakayama during the writing of this manuscript. Some of these results were presented at the 1990 meeting of the Association for Research in Vision and Ophthalmology in Sarasota, Florida (Verghese & Pelli, 1990). Most of these results appeared in Preeti Verghese's Ph.D. thesis (Verghese, 1990).